

SCIENTIFIC REPORTS



OPEN

Mycobacterium tuberculosis whole genome sequencing provides insights into the Manila strain and drug-resistance mutations in the Philippines

Jody E. Phelan¹, Dodge R. Lim^{2,3}, Satoshi Mitarai⁴, Paola Florez de Sessions⁵, Ma. Angelica A. Tujan⁶, Lorenzo T. Reyes^{2,3}, Inez Andrea P. Medado⁶, Alma G. Palparan^{2,3}, Ahmad Nazri Mohamed Naim⁵, Song Jie⁵, Edelwisa Segubre-Mercado⁶, Beatriz Simoes¹, Susana Campino¹, Julius C. Hafalla¹, Yoshiro Murase⁴, Yuta Morishige⁴, Martin L. Hibberd¹, Seiya Kato⁴, Ma. Cecilia G. Ama^{2,3} & Taane G. Clark^{1,7}

The Philippines has a high incidence of tuberculosis disease (TB), with an increasing prevalence of multidrug-resistant *Mycobacterium tuberculosis* (MDR-TB) strains making its control difficult. Although the *M. tuberculosis* “Manila” ancient lineage 1 strain-type is thought to be prevalent in the country, with evidence of export to others, little is known about the genetic diversity of circulating strains. By whole genome sequencing (WGS) 178 isolates from the Philippines National Drug Resistance Survey, we found the majority (143/178; 80.3%) belonged to the lineage 1 Manila clade, with the minority belonging to lineages 4 (European-American; $n = 33$) and 2 (East Asian; $n = 2$). A high proportion were found to be multidrug-resistant (34/178; 19.1%), established through highly concordant laboratory drug susceptibility testing and *in silico* prediction methods. Some MDR-TB isolates had near identical genomic variation, providing potential evidence of transmission. By placing the Philippine isolates within a phylogeny of global *M. tuberculosis* ($n > 17,000$), we established that they are genetically similar to those observed outside the country, including a clade of Manila-like strain-types in Thailand. An analysis of the phylogeny revealed a set of ~200 SNPs that are specific for the Manila strain-type, and a subset can be used within a molecular barcode. Sixty-eight mutations known to be associated with 10 anti-TB drug resistance were identified in the Philippine strains, and all have been observed in other populations. Whilst nine putative streptomycin resistance conferring markers in *gid* (8) and *rrs* (1) genes appear to be novel and with functional consequences. Overall, this study provides an important baseline characterisation of *M. tuberculosis* genetic diversity for the Philippines, and will fill a gap in global datasets and aid the development of a nation-wide database for epidemiological studies and clinical decision making. Further, by establishing a molecular barcode for detecting Manila strains it will assist with the design of diagnostic tools for disease control activities.

¹Infection Biology Department, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK. ²National Tuberculosis Reference Laboratory, Research Institute for Tropical Medicine, Muntinlupa City, Philippines. ³TB Study Group, Research Institute for Tropical Medicine, Muntinlupa City, Philippines. ⁴Department of Mycobacterium Reference and Research, Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Tokyo, Japan. ⁵Genome Institute of Singapore, 60 Biopolis St, Biopolis, Singapore. ⁶Molecular Biology Laboratory, Research Institute for Tropical Medicine, Muntinlupa City, Philippines. ⁷Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK. Jody E. Phelan and Dodge R. Lim contributed equally. Ma. Cecilia G. Ama and Taane G. Clark jointly supervised this work. Correspondence and requests for materials should be addressed to T.G.C. (email: Taane.clark@lshtm.ac.uk)

Tuberculosis disease (TB) is a chronic respiratory infection, caused by *Mycobacterium tuberculosis*. The Philippines is a high-burden country for TB, with over 300,000 cases and 25,000 deaths in 2017 alone¹. However, recent findings suggest that over a million Filipinos are actively infected with the disease, placing the Philippines burden second to China¹. Furthermore, the country has an increasing HIV prevalence and a high burden for multidrug-resistance (MDR-TB, resistance to isoniazid and rifampicin (RIF) treatments) that pose serious challenges for effective control¹. The Philippines has adopted the World Health Organisation (WHO) TB - “Directly Observed Treatment, Short Course” (TB-DOTS) control strategy. It has been recognised that there is an urgent need to detect and treat more cases through an expansion of GeneXpert (Cepheid, Sunnyvale, CA) facilities for better access and for use as a primary diagnostic tool¹. In particular, there is a gap among those expected to have MDR-TB (2% new, 21% retreatment cases) and those detected and subsequently put on treatment. Furthermore, treatment outcomes among those treated show poor success rates, thereby putting the country at risk to the further spread of MDR-TB and potentially extensively drug-resistant strains.

Despite TB being a serious problem in the Philippines, genomic data for local *M. tuberculosis* strains is lacking. Previous studies have identified ‘ancient’ (lineage 1) and ‘modern’ lineages (2 and 4) as being present in the population² with the members of the EA12-Manila clade known to be highly associated with the Filipino population^{3,4}. More isolates are needed to accurately infer the phylogeographic distribution of *M. tuberculosis* strains in the country, and take advantage of new genomic tools that allow for whole genome sequencing (WGS). WGS data can be used to profile the *M. tuberculosis* for drug resistance^{5,6}, characterise the ancient and modern lineages and different virulence strain-types⁷, and establish who may have transmitted to whom and thus allow targeted resources to hotspot areas to reduce transmission⁸; all made possible through advances in health informatics⁵. Further, WGS opportunities are set to revolutionise the diagnosis and clinical patient management of TB. For example, Public Health England (UK) now uses pathogen genetic characterisation as a clinical standard of care in TB management, and an increasing number of countries worldwide are seeking to adopt this as part of clinical care. However, to be effective these approaches need to be applied in countries with high TB burdens and coordinated with established control programmes.

In the Philippines, a nationwide drug resistance survey (DRS) occurs quinquennially (years 2007, 2012, 2017) and there is active surveillance of MDR-TB, including monitoring of fluoroquinolone and second line injectables resistance among TB patients, which involves the sequencing of drug resistance loci. This involves characterising single nucleotide polymorphisms (SNPs) and indel markers of drug resistance, especially in genes coding for drug-targets or -converting enzymes (e.g. *katG*, *inhA*, *rpoB*, *pncA*, *embB*, *rrs*, *gyrA*, *gyrB* genes). Limited sets of high frequency markers have been included in diagnostic tools (e.g. Xpert MTB/RIF and line probe assays for MDR-TB) for identification in *M. tuberculosis* DNA (approximate genome size 4.4 Mb). However, these approaches use less than 1% of the genome, and WGS approaches that identify SNPs and other variation provide far greater resolution for profiling *M. tuberculosis* for drug resistance, strain-types and detecting outbreaks, but also inform new mutations and mechanisms through genome-wide analyses^{5,9}.

To date, there have been few WGS studies in the Philippines, and here we present the results of sequencing 178 isolates recently collected in the 2012 DRS. We explore the genomic diversity of predominant Manila strain and develop a molecular barcode for surveillance applications. We show that the strain-types and majority of drug resistance mutations identified are similar to those found in global populations, making the use of existing mutation databases robust to detect resistance.

Results and Discussion

Clinical isolates and phylogeny. A total of 178 *M. tuberculosis* isolates were sourced from the nationwide “cluster-based”, cross-sectionally designed TB DRS (2012), which was implemented with cross re-checking standards according to the WHO guidelines. The isolates were randomly selected to represent the different geographic regions (“clusters”) in the Philippines. (Fig. 1A; Table 1). The isolates underwent culture, drug susceptibility testing (DST) across 8 drugs (Table 1), and WGS. The DST results revealed 97 (54.5%) isolates as pan-susceptible, 34 (19.1%) MDR-TB (resistance to both rifampicin and isoniazid) and 47 (26.4%) non-MDR-TB drug-resistant (resistance to at least one drug but not to both rifampicin and isoniazid). No XDR-TB isolates were found, confirming earlier reports of low levels of fluoroquinolone resistance¹⁰. Mapping of the raw sequence data led to high average genome-wide coverage across the samples (median: 100.5-fold; range: 52- to 2,263-fold). Across the clinical isolates, 17,522 unique SNPs were identified, and a high proportion (50%) were observed in single isolates. Isolates were classified into lineages 1 (n = 143; 80.3%), 4 (n = 33; 18.5%) and 2 (n = 2; 1.1%), and as expected these form clusters on the phylogeny constructed using all SNP sites (Fig. 1B). By inferring spoligotypes from the WGS data, the EA12-Manila spoligotype (lineage 1.2.1)³ was found to have the highest frequency (n = 143, 80.3%) among lineage 1 strains (Table 1).

Evidence of transmission. Potential transmission clusters were found by calculating the pairwise SNP distance between isolates and using an established difference cut-off of 10 SNPs or less⁸. Fifteen clusters were found, with a maximum size of three isolates. Of these clusters, nine contained MDR-TB strains, ten belonged to lineage 1, and five to lineage 4 (Supplementary Fig. 2). In order to establish whether particular strains are circulating locally or nation-wide, the phylogeographical distribution of the 178 strains was analysed (Supplementary Fig. 3A). There was no association of island group (Luzon, Mindanao and Visayas) with phylogenetic placement (Slatkin-Maddison P = 0.87). Using the closest neighbour for each strain, a weak correlation between genetic distance (SNP differences) and geographic distance was observed (Spearman's correlation = 0.36) (Supplementary Fig. 3B).

The Manila strain. This collection represents the largest number of Manila strains sequenced from the Philippines with WGS (n = 143). To determine if the Philippine strains formed their own monophyletic clade compared to other strains, we reconstructed a phylogenetic tree that also included lineage 1.2.1 strains with

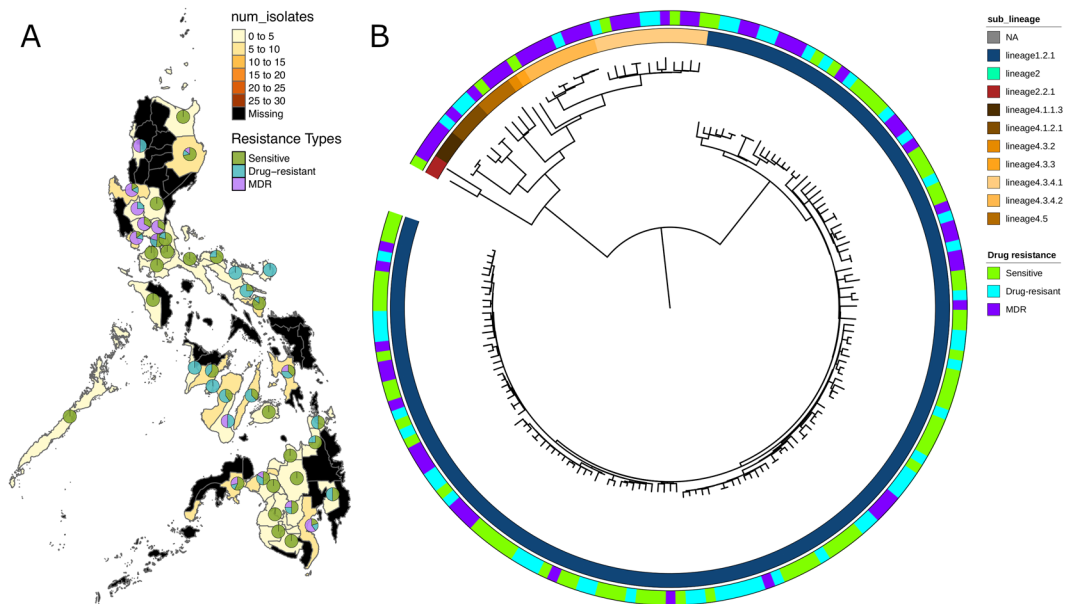


Figure 1. The 178 *M. tuberculosis* isolates: (A) Map of the Philippines annotated with the source and drug-resistance of the isolates; (B) A phylogenetic tree constructed using 17,522 SNPs.

publicly available sequences ($n = 408$). The Philippine isolates did not form a monophyletic group within the tree and were interleaved with strains from Thailand, UK, Saudi Arabia, India, and Vietnam among others (Supplementary Fig. 4). Most of these were sporadically scattered throughout the tree, however a large monophyletic clade with 114 isolates formed comprising of predominantly Thai strains. To determine the geographic origin of the most recent common ancestor (MRCA) of this clade, we performed ancestral state reconstruction using stochastic character mapping. The most likely location of the Thai clade MRCA was estimated to be from the Philippines (posterior probability = 0.79). Based on this, a likely scenario is that lineage 1.2.1 strains in Thailand were imported from the Philippines, after which they diversified. *In-silico* spoligotyping of this clade revealed a very different strain-type from the Manila spoligotype, with the loss of spacers 7–18 and 21–24, corresponding to the previously described EAI2-Nonthaburi spoligotype¹¹. A comparison of allele frequency differences between the Manila ($N = 408$) and non-Manila strains in neighbouring clades ($N = 1,636$) revealed 197 SNPs which were specific to the Manila clade. Of these, 180 were found in coding regions and 115 led to amino acid changes, including one premature stop codon in *whiB5*, which has been described before¹² (Supplementary Data 1).

Mutations underlying drug-resistance. Resistance profiles were predicted *in silico* using an established library of mutations (<https://github.com/jodyphelan/tbdb>), and were compared to DST results. Assuming DST as the gold standard, the sensitivity and specificity was calculated for rifampicin, isoniazid, ethambutol and streptomycin, where all had at least 10 isolates with demonstrated phenotypic resistance. The sensitivity of prediction was high (>90%) across the drugs except streptomycin. The specificity was high (>95%) for all drugs (Supplementary Table 1). The relative low sensitivity of the streptomycin predictions was investigated and revealed nine novel polymorphisms in the *gid* (8 mutations) and *rrs* (1 mutation) genes in 12 false negative isolates, which could potentially confer drug resistance. To investigate these mutations, we looked at their frequency in a large dataset of ~17,000 publicly available strains. Of the nine mutations, six were found in the global dataset and only one of these was found to be a lineage-specific marker (*gid* L16R) with the other five displaying characteristics of convergent evolution. The functional consequences of the mutations in *gid* were investigated, and it was established that two (W45*, S181*) led to premature stop codons, two (CA49C, AC194A) led to frameshifts, and three (G71R, V77G and D85G) were within or close to the S-adenosyl-L-methionine (SAM)-binding motif. The SAM-binding motif is an important functional domain required by *gid* to methylate the G527 on the 530 loop of 16S rRNA, which is thought to be the binding location of streptomycin¹³ (Supplementary Fig. 5). These mutations could represent resistance conferring mutations and should be investigated further. Overall, 68 known drug resistance-related mutations were found across the 178 samples, and they varied in frequency (range: 1 to 63 isolates) (Table 2). Although no phenotypic resistance to fluoroquinolones was found, one isolate had the N499K mutation in the *gyrB* gene. This mutation has been reported to be rare in clinical strains¹⁴ and is associated with resistance to moxifloxacin but not to ofloxacin or levofloxacin¹⁵. Predicted resistance to (non-streptomycin) aminoglycosides was also rare with only a single isolate presenting a resistance conferring mutation in *rrs* (A1401G).

Summary. WGS is being used as a tool for epidemiological investigations and to assist TB clinical and control program decision making. However, most clinical applications of WGS are taking place in developed countries. We present the first WGS data from the nation-wide TB drug resistance survey in the Philippines. Most isolates were from the EAI2-Manila clade, which would make this the largest WGS study of this spoligotype, but is also

Characteristic	N	%
<i>Location*</i>		
Luzon	105	59.0
Visayas	37	20.8
Mindanao	36	20.2
<i>Lineages</i>		
1	143	80.3
2	2	1.1
4	33	18.5
<i>Drug resistance**</i>		
Rifampicin	38	24.7
Isoniazid	64	46.0
Ethambutol	10	6.5
Streptomycin	25	16.3
Amikacin	2	1.1
Capreomycin	2	1.2
Kanamycin	8	4.5
Pan-susceptible	97	54.5
MDR-TB	34	21.5
XDR-TB	0	0

Table 1. *Mycobacterium tuberculosis* samples (N = 178). *Island of collection for sequenced isolates. **The number resistant to the drug and percentage of tested isolates resistant; the denominator is the number of isolates for each drug with successfully performed susceptibility tests. The exception to this is the proportion of pan-susceptible which was calculated using the total number of samples. MDR-TB = multi-drug resistant; XDR-TB = extensively-drug resistant; there was no levofloxacin drug resistance.

indicative that this clade is common. For comparison, a recent report of WGS of 10 Philippines samples included 7 EAI2-Manila, 1 Beijing and 2 Euro-American strains². In our WGS data, there was high SNP variability between isolates in the EAI2-Manila clade, and we identified 197 SNPs specific to the EAI2-Manila clade which could form the basis of a SNP barcode. There was some evidence of near identical isolates, which is indicative of transmission events. This observation is surprising as the samples were chosen pseudo-randomly from a large collection, but it may be indicative of many large transmission chains in the population. Future large-scale sequencing of isolates across the surveys is required to understand transmission patterns and dynamics. There is no evidence of a correlation between genetic and geographic distance, but this may reflect the within-country movement and non-localised transmission of strains in the Philippines. Finally, we performed *in-silico* drug resistance profiling and compared this to phenotypic DST results. We found a number of different mutations conferring resistance to first-line anti-TB drugs, and report novel potential resistance conferring mutations, which can be investigated experimentally. Further, the mutations observed are similar to those reported in other drug resistant *M. tuberculosis* from the Philippines² and globally. Overall, our work reveals the utility of WGS for the prediction of drug resistance in the Philippines setting. The whole genome data generated will serve as a reference for the surveillance of TB in the Philippines and the wider region.

Materials and Methods

DNA extraction and sequencing. A total of 178 *M. tuberculosis* short-term cultured isolates from sputum samples collected from the second national DRS on TB in the Philippines (2012) were pseudo-randomly selected for the study, where selection was informed by the WHO cross re-checking standard. This study was given authorization by the Institutional Review Board of the Research Institute for Tropical Medicine in the Philippines (ID No. RITM-IRB 2017-05). The clinical samples were complemented by two replicates of the H37Rv reference strain. Drug susceptibility testing was performed for rifampicin, isoniazid, ethambutol, streptomycin, amikacin, kanamycin, capreomycin and levofloxacin at the National Tuberculosis Research Laboratory (NTRL) (Supplementary Table 1). For the H37Rv and 14 clinical isolates, inactivation of culture isolates was performed in a BSL-3 facility at the NTRL, and DNA extraction at the Molecular Biology Laboratory at the Research Institute for Tropical Medicine (RITM). Total genomic DNA was extracted using a phenol-chloroform extraction procedure. DNA pellets were washed with 70% ethanol and resuspended in 100ul Buffer EB (QIAGEN Inc., Germantown, MD, USA). DNA extract concentration and quality were measured using a Qubit fluorometer (Life Technologies Holdings Pte Ltd, Singapore) and were visualized on a 1% agarose gel. Library preparation of the DNA samples was performed using QIAseq FX DNA library kit, following the manufacturer's protocol. Agilent High Sensitivity DNA Kit (QIAGEN Inc., Germantown, MD, USA) was used to check for correct DNA fragment size after adapter ligation. The concentration and size of each DNA library was measured using Agilent Technologies DNA 1000 assay kit on Agilent Technologies 2100 Bioanalyzer. The H37Rv prepared DNA libraries were sequenced using Illumina HiSeq. 4000 Next Generation Sequencer (2 × 151 base pair reads) at the Genome Institute Singapore. Similarly, for the other 164 clinical strains, DNA extraction and purification were performed in the BSL-3 laboratory in the Research Institute of Tuberculosis (RIT) using Isoplant II (WAKO chemicals,

Drug	Gene	Mutation	Freq. (%)	Global freq* (%)
Isoniazid	<i>katG</i>	S315T	63	4057/5423 (74.8)
Rifampicin	<i>rpoB</i>	S450L	37	2968/4618 (64.3)
Isoniazid/Ethionamide	<i>inhA promoter</i>	C-15T	35	872/5423 (16.1)
Ethambutol	<i>embB</i>	M306I	15	564/2662 (21.2)
Streptomycin	<i>rrs</i>	A514C	11	128/1408 (9.1)
Streptomycin	<i>rpsL</i>	K88R	7	103/1408 (7.3)
Rifampicin	<i>rpoB</i>	H445Y	6	230/4618 (5.0)
Isoniazid	<i>inhA</i>	I21V	5	15/5423 (0.3)
Ethambutol	<i>embB</i>	M306V	4	839/2662 (31.5)
Isoniazid/Ethionamide	<i>inhA promoter</i>	T-8C	4	89/5423 (1.6)
Ethambutol	<i>embB</i>	Q497K	3	20/2662 (0.8)
Isoniazid/Ethionamide	<i>inhA promoter</i>	G-17T	3	72/5423 (1.3)
Isoniazid/Ethionamide	<i>inhA</i>	I21T	3	32/5423 (0.6)
Streptomycin	<i>gid</i>	102del	3	12/1408 (0.9)
Streptomycin	<i>gid</i>	351del	3	6/1408 (0.4)
Streptomycin	<i>rpsL</i>	K43R	3	642/1408 (45.6)
Ethambutol	<i>embA</i>	D4N	2	3/2662 (0.1)
Ethambutol	<i>embB</i>	D328G	2	16/2662 (0.6)
Ethambutol	<i>embB</i>	G406D	2	75/2662 (2.8)
Isoniazid/Ethionamide	<i>inhA</i>	I194T	2	88/5423 (1.6)
Pyrazinamide	<i>pncA</i>	D12A	2	11/1946 (0.6)
Pyrazinamide	<i>pncA</i>	G24D	2	3/1946 (0.2)
Pyrazinamide	<i>pncA</i>	I133T	2	24/1946 (1.2)
Rifampicin	<i>rpoB</i>	1297_1298insTTC	2	11/4618 (0.2)
Rifampicin	<i>rpoB</i>	1295_1303del	2	5/4618 (0.1)
Rifampicin	<i>rpoB</i>	D435F	2	23/4618 (0.5)
Rifampicin	<i>rpoB</i>	S441L	2	17/4618 (0.4)
Streptomycin	<i>gid</i>	386del	2	2/1408 (0.1)
Streptomycin	<i>rrs</i>	C517T	2	57/1408 (4.1)

Table 2. Drug resistance mutations with frequency greater than one. *Based on the 17k dataset.

Japan). The clinical strains were sequenced on the Illumina MiSeq platform (350 forward and 250 reverse pair ends) at the RIT. All raw sequencing data is available (Supplementary Data 2).

Bioinformatic analyses. Sequence reads were inspected using fastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) as a primary assessment of data quality. The reads were trimmed using trimmomatic¹⁶ (v0.38) to remove low quality sequences, and then mapped against the H37Rv reference genome (AL123456) using BWA (v0.7.17)¹⁷. SNPs were called using the BCF/VCF tool suite (v1.8)¹⁸ in regions where at least 10 reads were present. SNPs were removed from non-unique regions of the genome (e.g. *ppe* genes). SNPs were converted into a FASTA format alignment, which was used as input to RAxML (v8.0.0)¹⁹ to reconstruct the phylogeny. Drug resistance profiles and lineages were predicted *in-silico* using TBProfiler (v2.0). SpolPred (v1)²⁰ was used to *in-silico* predict spoligotypes. As expected the resolution from using spoligotypes alone did not have the power to differentiate distinct WGS-based clusters found on the phylogenetic tree (Supplementary Fig. 1). The Philippine isolates were compared to others in a global set⁶ with the same lineage (see Supplementary Data 3) and F_{ST} values were calculated for each SNP using vcftools (-weir-fst-pop) to identify Manila-strain specific markers. A Slatkin-Maddison test was applied to assess a geographic-phylogenetic association and was performed in R using the *prmac/slatkin.maddison* library.

Data Availability

Previously published and newly generated data can be found on the ENA using the Run accession codes in Supplementary Data 2. The newly generated data can be found under the ENA study accession number ERP110368.

References

1. Global tuberculosis report (www.who.int/tb/publications/global_report/en) (2018).
2. Roa M. B. *et al.* Whole genome sequencing and single nucleotide polymorphisms in multi-drug resistant clinical isolates of *Mycobacterium tuberculosis* from the Philippines. *J Glob Antimicrob Resist*, <https://doi.org/10.1016/j.jgar.2018.08.009> (2018).
3. Douglas, J. T. *et al.* Characterization of the Manila family of *Mycobacterium tuberculosis*. *J Clin Microbiol* **41**, 2723–6, <https://doi.org/10.1128/JCM.41.6.2723-2726.2003> (2003).
4. Montoya, J. C., Murase, Y., Ang, C., Solon, J. & Ohkado, A. A molecular epidemiologic analysis of *Mycobacterium tuberculosis* among Filipino patients in a suburban community in the Philippines. *Kekkaku* **88**, 543–52 (2013).

5. Coll, F. *et al.* Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med* 7, 51, <https://doi.org/10.1186/s13073-015-0164-0> (2015).
6. Coll, F. *et al.* Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. *Nat Genet* 50, 307–16, <https://doi.org/10.1038/s41588-017-0029-0> (2018).
7. Coll, F. *et al.* A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun* 5, 4812, <https://doi.org/10.1038/ncomms5812> (2014).
8. Guerra-Assunção, J. *et al.* Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *Elife* 4, <https://doi.org/10.7554/eLife.05166> (2015).
9. Phelan, J. *et al.* *Mycobacterium tuberculosis* whole genome sequencing and protein structure modelling provides insights into anti-tuberculosis drug resistance. *BMC Med* 14, 31, <https://doi.org/10.1186/s12916-016-0575-9> (2016).
10. Lim, D. R. *et al.* Low prevalence of fluoroquinolone resistance among patients with tuberculosis in the Philippines: results of a national survey. *Eur Respir J* 51, 1702571, <https://doi.org/10.1183/13993003.02571-2017> (2018).
11. Coker, O. O. *et al.* Genetic signatures of *Mycobacterium tuberculosis* Nonthaburi genotype revealed by whole genome analysis of isolates from tuberculous meningitis patients in Thailand. *PeerJ* 4, e1905, <https://doi.org/10.7717/peerj.1905> (2016).
12. Wan, X. *et al.* Genomic analyses of the ancestral Manila family of *Mycobacterium tuberculosis*. *PLoS One* 12, e0175330, <https://doi.org/10.1371/journal.pone.0175330> (2017).
13. Okamoto, S. *et al.* Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. *Mol Microbiol* 63, 1096–106, <https://doi.org/10.1111/j.1365-2958.2006.05585.x> (2007).
14. Maruri, F. *et al.* A systematic review of gyrase mutations associated with fluoroquinolone-resistant *Mycobacterium tuberculosis* and a proposed gyrase numbering system. *J Antimicrob Chemother* 67, 819–31, <https://doi.org/10.1093/jac/dkr566> (2012).
15. Malik, S., Willby, M., Sikes, D., Tsodikov, O. V. & Posey, J. E. New Insights into Fluoroquinolone Resistance in *Mycobacterium tuberculosis*: Functional Genetic Analysis of *gyrA* and *gyrB* Mutations. *PLoS One* 7, e39754, <https://doi.org/10.1371/journal.pone.0039754> (2012).
16. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–20, <https://doi.org/10.1093/bioinformatics/btu170> (2014).
17. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (2013).
18. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–93, <https://doi.org/10.1093/bioinformatics/btr509> (2011).
19. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–3, <https://doi.org/10.1093/bioinformatics/btu033> (2014).
20. Coll, F. *et al.* SpolPred: rapid and accurate prediction of *Mycobacterium tuberculosis* spoligotypes from short genomic sequences. *Bioinformatics* 28, 2991–3, <https://doi.org/10.1093/bioinformatics/bts544> (2012).

Acknowledgements

The project is funded by the Philippine General Appropriations Act, Japan Agency for Medical Research and Development (AMED) (JP16fk0108204), Research Institute of Tuberculosis (RIT), Genome Institute of Singapore, an MRC UK – PCHRD - Newton Agham award (Grant no. MR/R025576/1), and a British Council – CHED – Newton Institutional links grant (Ref. 261868591). TGC is funded by the Medical Research Council UK (Grant no. MR/M01360X/1, MR/N010469/1, MR/R025576/1, and MR/R020973/1) and BBSRC (Grant no. BB/R013063/1). SC is funded by Medical Research Council UK grants (MR/M01360X/1, MR/R025576/1, and MR/R020973/1) and BBSRC (Grant no. BB/R013063/1). We gratefully acknowledge Dr. Socorro P. Lupisan, Dr. Amado O. Tandoc III, Dr. Ramon P. Basilio, Catherine PL Sacopon, Michellin Baje, Angeli Borbe-Reyes, and Niña Hunat for their role in the National Drug Resistance Survey on TB and ongoing clinical studies in the RITM; the Scientific Computing Group for data management and compute infrastructure at the Genome Institute of Singapore for their help. The MRC eMedLab computing resource was used for the bioinformatics and statistical analysis.

Author Contributions

T.G.C., D.R.L., S.M., S.K. and M.C.G.A. conceived and directed the project. D.R.L., P.F.D.S., Y.Mu., Y.Mo., S.M. and M.L.H. coordinated sample collection. D.R.L., L.T.R., M.A.A.T., I.A.P.M., A.G.P., A.N.M.N., Y.Mu., Y.Mo., S.J. and E.S.-M. undertook sample collection, processing, DNA extraction, library building and sensitivity testing. P.F.D.S., J.C.H., Y.Mu., Y.Mo., S.M. and M.L.H. coordinated sequencing. J.E.P. and B.S. performed bioinformatic and statistical analyses under the supervision of S.C., M.L.H. and T.G.C. J.E.P., D.R.L., M.L.H. and T.G.C. interpreted results. J.E.P., D.R.L. and T.G.C. wrote the first draft of the manuscript. All authors commented and edited on various versions of the draft manuscript and approved the final manuscript. J.E.P., D.R.L., M.C.G.A. and T.G.C. compiled the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-45566-5>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019